Chapter 4

The “neglected” soil virome – potential role and impact

Akbar Adjie Pratama and Jan Dirk van Elsas

Published in *Trends Microbiol*. (2018) 1533 1–14
doi: 10.1016/j.tim.2017.12.004

Abstract

Bacteriophages are among the most abundant and diverse biological units in the biosphere. They have contributed to our understanding of the central dogma of biology and have been instrumental in the evolutionary success of bacterial pathogens. In contrast to our current understanding of marine viral communities, the soil virome and its function in terrestrial ecosystems has remained relatively understudied. Here, we examine, in a comparative fashion, the knowledge gathered from studies performed in soil versus marine settings. We address the information with respect to the abundance, diversity, ecological significance and effects of, in particular, bacteriophages on their host's evolutionary trajectories. We also identify the main challenges that soil virology faces and the studies that are required to accompany the current developments in marine settings.

Keywords: Virus community, soil, ecology, evolution

Why study soil viromes?

Soil plays a pivotal role in the functioning of the global biome as it provides a multitude of different ecosystem services, including key processes of the global biogeochemical cycles (cycling of C and N). Moreover, soils allow agricultural use in a heterogeneous habitat that supports a high diversity of microbes (Garbeva et al., 2004). The soil microbiomes encompass a myriad of diverse organisms (including bacteria, archaea, fungi, protozoa, and nematodes, next to the viruses that infect them). The interactions of these organisms with their respective viruses play essential roles in the aforementioned soil functions. Given the fact that bacteria are by far the most numerous organisms in soil, key interest has developed in their abundance and diversity in soil systems. This pertains to agricultural (Trivedi et al., 2016), forest (Brabcová et al., 2016), plant-root-associated (i.e. rhizosphere) (Yan et al., 2017), salt marsh (Dini-Andreote et al., 2014), permafrost (Jansson and Taş, 2014) and artic soils (Blaud et al., 2015). Thus, whereas the organismal members of soil bacteriomes are increasingly addressed in key soil studies, there is a paucity of knowledge with respect to the role of their associated **virome** (see Glossary).

Highlights

Studies of the soil virome have been largely neglected, whereas viromics studies have yielded important concepts in marine settings.

The heterogeneity of soil provides an environment where microbes act and interact in parallel, yet in close proximity, pursuing parallel evolutionary paths. The soil virome contributes to this multiple simultaneous parallel evolution scenario.

There are still major challenges in soil virology that need to be addressed to keep pace with its counterpart in marine virology. These include: (*i*) underutilized technologies, both experimental and computational, (*ii*) the lack of knowledge about phage life cycles and activities in soil, and (*iii*) the lack of spatiotemporally explicit sampling.

An improved understanding of the soil virome is essential to understand the extent of the role of viruses in the ecological functioning and evolution of microorganisms.

Recent metagenomics-based data have indicated that, at least in marine settings, viral abundances are often as high as, or higher than, bacterial ones (Hatfull, 2015; Paez-Espino et al., 2016). Among the identified viruses, **bacteriophages** (**phages**) in particular were found to be prevalent (Paez-Espino et al., 2016, 2017a; Roux et al., 2016a; Zablocki et al., 2014b). Since their discovery, about a century ago (independently

by Twort and d'Herelle (Salmond and Fineran, 2015), phages have had a key role in the development of the fundamentals of biology (for instance, establishing the central dogma of biology, i.e., DNA as the carrier of inherited information). Thus, the seminal work with phages, performed in the 1950s, has greatly advanced the development of key insights into their roles as population controllers and vectors of **horizontal gene transfer** (**HGT**) (Salmond and Fineran, 2015). More recent studies on *Escherichia coli* (Ohnishi et al., 2001) and *Mycobacterium* spp. (Pedulla et al., 2003) indeed revealed that bacteriophages have – over evolutionary time – played significant roles in shaping their host's genomes. To what extent we can extrapolate such findings to soil settings has, however, remained enigmatic. Clearly, the currently available advanced DNA sequencing technology will – in principle – enable a strong development in the soil virome area (Breitbart, 2012; Brum and Sullivan, 2015; Paez-Espino et al., 2016; Roux et al., 2016a; Suttle, 2005, 2007) provided key issues pertaining to soil settings can be resolved.

In this opinion article we examine the state-of-the-art of current environmental virome studies, with a focus on soil in comparison to marine systems. We acknowledge the broad scope of the virome in environmental settings, which includes the viruses of bacteria, Archaea, and a range of Eukarya (Ghabrial et al., 2015; Krupovic et al., 2011; Prangishvili et al., 2017). Considering the dominance of bacteria in soil ecosystems, in this article we place emphasis on the bacteriophage part of the soil virome. After a description of the selective forces exerted on the host organisms in marine versus soil settings, we examine our current understanding of the abundance, diversity, evolution, and putative role of their associated viruses. We also identify the current challenges in this area and suggest potential strategies to further foster our knowledge.

Marine versus soil habitats - implications for viral life

Marine habitats as 'homes' for microbiomes are characterized by processes such as (differential) mixing,particlebuild-upandsinking,andstratification.Moreover,vertical gradients with respect to light as an energy source, and oxygen as a terminal electron acceptor, drive the selection of organisms (Whittaker and Rynearson, 2017). Microorganisms living in marine settings often interact with suspended particles (as well as with other organisms) or are themselves suspended. In such states, they are subjected to the local conditions, under which (shifts in) the redox and nutritional environments are key factors. In contrast, fluctuations in local temperature and pH are thought to be moderate in the marine aquatic system, as a result of the buffering capacity of the water body (Zehr and Ward, 2002). These conditions of

mixing/connectivity and 'smoothening out' of shifts in temperature and pH form the background against which virus–host interactions shape up.

Soil habitats are intrinsically heterogeneous and diverse (Kuzyakov and Blagodatskaya, 2015). The spatial structure of soil, as seen in soil aggregates, creates multiple microhabitats and thus niches for microorganisms such as bacteria to develop. On top of the spatial heterogeneity resulting from soil structure, particular sites in soil, such as the zones of influence of plant roots, fungal hyphae, and/or earthworms, can – ephemerally – show enhanced nutritional status, consequently creating microbial activity hotspots (**Figure 4.1**, Key Figure). Thus, under water-unsaturated conditions, each site in soil, for example, an aggregate, may harbor different microbial communities, ranging from several thousands to several millions of individuals (Rillig et al., 2017; Vos et al., 2013). The consequent spatial isolation, and lack of connectivity, of the individual 'island' microbiomes within aggregates promotes parallel microbial evolution trajectories. Such parallel evolutionary events are bound to increase the local diversity within soil microbiomes, even within narrow taxonomic rankings (Rillig et al., 2017; Stefanic and Mandic-Mulec, 2009). However, soil incidentally becomes waterlogged and, consequently, the island microbiomes can, to some extent, mix and diminish the parallel evolutionary effect (Rillig et al., 2017). Moreover, the soil aggregates may also incidentally disintegrate, increasing the level of connectivity across aggregates, for example, through the water present in the soil. Here, we posit that the selection pressures inside a 'hot spot' soil aggregate, for example by accretion and subsequent depletion of nutrients, may promote phage activity, potentially contributing to the aforementioned evolutionary scenario. The continuous formation and disintegration of soil aggregates, and the timing thereof, may thus (indefinitely) affect the pools of, and fluxes in, soil microbiome diversity (Kuzyakov and Blagodatskaya, 2015; Rillig et al., 2017).

Approaches to the study of marine and soil viromes

The notion that bacteriophages can infect only a limited number of different bacterial hosts has long limited our understanding of their roles in systems that are amenable to experimentation (mainly by cultivation). However, with the current advanced molecular technologies, broader studies, across marine and soil settings, are possible (Breitbart, 2012; Brum and Sullivan, 2015). The early studies thus relied on enumeration methods, for example, via the double-agar-layer (DAL) method (Adams, 1959), or by TEM (transmission electron microscopy) and/or EFM (epifluorescence microscopy) (Ashelford et al., 2003; Swanson et al., 2009; Williamson et al., 2005). Indeed, EFM nowadays constitutes a routine technique for marine virome studies.

Glossary

Auxiliary metabolic genes (AMGs): metabolic genes, carried by viruses, that enhance host fitness.

Bacteriophages (phages): viruses that infect, and reproduce in, bacteria.

Coevolution: evolutionary process defined as the reciprocal adaptation between interacting partners.

Dissolved organic matter (DOM): organic matter smaller than 0.45 mm, derived from decomposition of organic matter derived from plants, animals, and bacteria.

Horizontal gene transfer (HGT): lateral transfer of genetic information, that is, not from parent to offspring.

Kill-the-winner (KtW): lytic virus regulates bacterial populations by eradicating the most abundant bacterial strain.

Mobile genetic elements: mobile elements consisting of DNA that are able to move among bacteria, for example, plasmids, phages, transposons, group I introns and group II introns.

Moron gene(s): virus gene(s) not required for virus replication; quite often, these genes offer fitness advantages to host cells.

Particulate organic matter (POM): organic matter larger than 0.7 mm, derived from decomposition of organic matter, including plants and bacteria.

Piggy-back-the-winner (PtW): lysogenic virus–host interaction in highly dense host populations.

Prophage: the integrated version of a bacterial virus in the bacterial host genome; some prophages are able to switch to a lytic cycle under certain circumstances.

Temperate phages: bacterial viruses that can integrate into, and stabilize in, a bacterial genome (or be maintained extrachromosomally) and, upon receiving a cue, can switch to a lytic cycle, excise, and propagate.

Viral shunt process: the process in which a lytic virus mediates the recycling of dissolved organic matter.

Virome: the entire virus community.

Viromics: see Virus communities'metagenomics (below).

Virulence factor: a molecule (e.g., Shiga toxins *stx*-1 and *stx*-2) that enables the successful increase in pathogenicity of a pathogen.

Virulent virus: a virus with a lifecycle in which lysis of host cells is key in propagation.

Virus communities' metagenomics (viromics): the study of an entire virus community through metagenomics.

Virus-like particle (VLP): a structure that resembles a virus, that is, a virus capsid or virus tails.

Virus-to-bacterium ratio (VBR): the relative abundance of viruses compared to bacteria.

It is based on the fact that fluorescent dyes, such as SYBR green and SYBR gold, have the propensity to strongly bind to double-stranded DNA (dsDNA) and RNA, allowing extremely sensitive staining of virions (Patel et al., 2007). EFM has also been applied to soil samples; however, great obstacles to its successful use (i.e., due to aspecific interaction with soil particles) have been identified (Ashelford et al., 2003). Hence, researchers have preferred to use TEM [next to plaque-forming unit (PFU) counts] for enumeration of soil phages (Swanson et al., 2009; Williamson et al., 2013) and assessments of soil viral diversity (Williamson et al., 2005). However, TEM is laborious and it cannot – albeit with some exceptions – give meaningful classifications solely on the basis of virion morphologies. For instance, the observable parameters (e.g., virion head and tail sizes and shape) may be indistinguishable within and between viral families (Thurber et al., 2017). Thus, we advocate the use of meta-omics methods to advance the field. Whereas such methods have shown their merits in marine viromics studies, they pose specific problems for soil. Clearly, the heterogeneity of soil makes any sample taken appear 'unique' in terms of its microbiome and virome. Adequate sampling, thus, is a highly critical issue in soil studies. Indeed, in the light of the parallel evolution trajectories depicted in the foregoing, microscale sampling may be required, which poses specific problems of sample numbers and scale. Other issues pertain to the adherence of viral particles, as well as microbiome or virome nucleic acids, to the soil.

We advocate the further development of studies based on pre-extracted virions or nucleic acids from soil. However, the absence of a universal molecular marker for viruses or phages, next to the classical approach to viral classification based on morphology (capsid sizes and shapes) has hitherto limited the rapid development of environmental viromics. Moreover, the diversity of viral nucleic acid types [single-stranded RNA (ssRNA), ssDNA, double-stranded RNA (dsRNA), and dsDNA] has posed problems (Simmonds et al., 2017). The Bacterial and Archaeal Viruses Subcommittee (BAVS) of the International Committee on Taxonomy of Viruses (ICTV) has pointed the way forward in viral classification; this may also be useful in environmental viromics. Thus, the recommended approaches to classifying viruses are based on genome-to-genome analyses of individual viral isolates (Hatfull, 2015), placing a focus on viral signature genes [e.g., for phages, portal or major capsid genes (Myoviridae family),

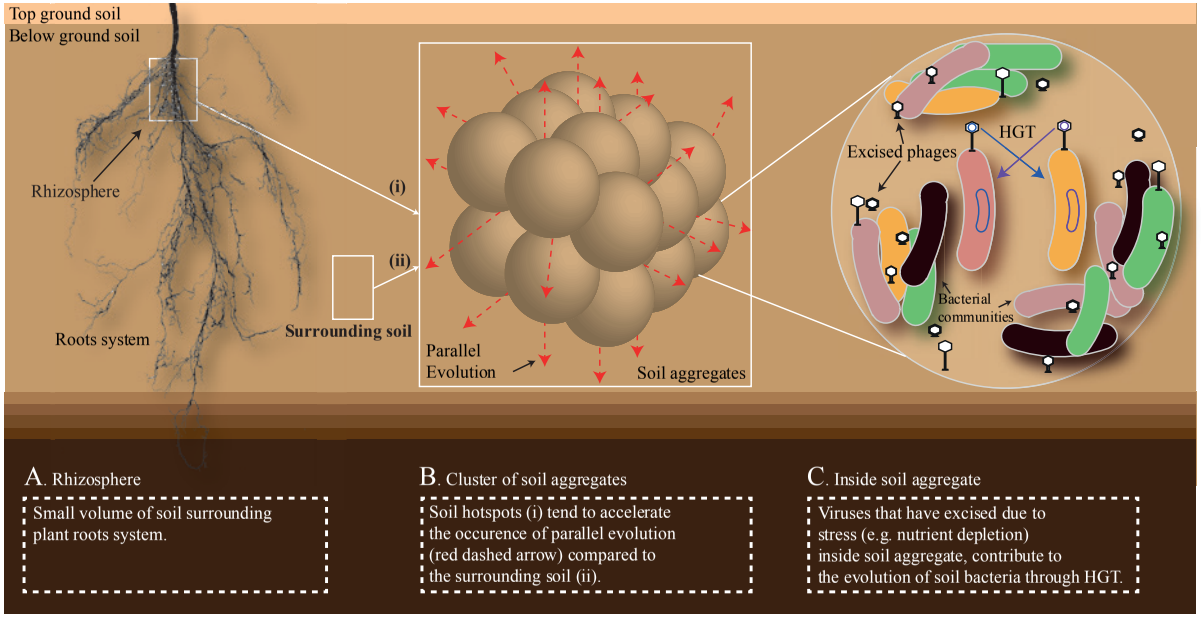

Figure 4.1. Proposed scenario of the evolutionary contribution of the virome to the soil microbiome. Most activity in the soil ecosystem can be observed in microhabitats denoted activity hotspots, which can be heterogeneous as a result of soil development. The factors that determine the formation of soil hotspots include: redox potential, organic matter level and type, nutrient content, microbial biomass, composition of the local microbiome, and local enzyme activities. The consequence of microbial activity is often an increase in microbial abundance (and possibly diversity (Kuzyakov and Blagodatskaya, 2015). Recognized hot spots are (*i*) the rhizosphere, (*ii*) the detritusphere, (*iii*) biopores, (*iv*) aggregate surfaces, and (*v*) the mycosphere. All of these hotspots have roles in the functioning of the soil (e.g., decomposition, mineralization of soil organic matter, nutrient cycling, and C turnover), as they increase microbial processes and interactions (Kuzyakov and Blagodatskaya, 2015). The soil aggregates in the hotspots offer spatial isolation for bacteria to pursue their own evolution (Rillig et al., 2017; Vos et al., 2013). The figure illustrates (**A**) the rhizosphere; (**B**) a cluster of soil aggregates that exists in both rhizosphere and surrounding soil [indicated by white boxes and arrows (i and ii)]; parallel evolution (indicated by red dashed arrow); and (**C**) the conditions inside a soil aggregate (e.g., nutrient depletion) which can induce the excision of phages. Thus, phages are likely to participate in bacterial evolutionary trajectories through induction and horizontal gene transfer (HGT, indicated by blue and indigo arrows).

photosynthesis-related genes *psbA* and *psbD*, and T7-like DNA polymerase gene *polA*] (Adriaenssens and Cowan, 2014). Classifications based on protein sequences (Rohwer and Edwards, 2002) and phage network cluster analyses (Lima-Mendez et al., 2008) have also been proposed, and, additionally, computational virome clustering based on reference-independent methods (Hurwitz et al., 2016) and computational frameworks (which relies on both average nucleotide identity and total alignment fraction for pairwise comparisons of viral sequences) has come into play (Paez-Espino et al., 2017b). Progress in the area may be further fostered by the use of either defined gene sequences as proxies for viral presence and/or abundance, or of computationally defined hidden Markov model (HMM)-like algorithms. On the basis of such approaches, the effects of, for instance, environmental shifts on viral dynamics in the system, might be observed. However, although promising, bottlenecks still persist in the current computational methods, as we discuss later.

Viral abundance - marine versus soil ecosystems

Direct **VLP** (**virus-like particle**) counts based on TEM and EFM have revealed that there are commonly 10^6 – 10^7 viruses ml^{-1} of water in most pelagic marine settings. Similar results (of about 10^6 VLP ml^{-1}) have been reported in both coastal and offshore waters (Suttle, 2007; Thurber et al., 2017). On the basis of morphological criteria, the contention that there is a high viral diversity was confirmed, the majority of viruses being tailed phages (i.e., Myoviridae, Podoviridae, Siphoviridae) (Breitbart, 2012; Breitbart and Rohwer, 2005; Suttle, 2005, 2007). Moreover, surveys in which virome metagenomics (viromics) was used (i.e., the Tara Ocean and Malaspina Expeditions, the US microbial observatory at the San Pedro Ocean time series, the Pacific Ocean virome and the Bermuda Atlantic time series studies) all provided essential spatiotemporally explicit data on the abundance and diversities of the viromes studied (Brum and Sullivan, 2015). Such advanced viromics studies have greatly accelerated the pace of marine virology (Brum and Sullivan, 2015; Roux et al., 2016a), with current datasets containing about 140,134 viral contigs (IMG/VR)[i] (Paez-Espino et al., 2017a). However, given the biases against ssDNA and RNA viruses in metagenomics pipelines, the environmental abundance and diversity of these viruses is currently unclear (Brum and Sullivan, 2015).

A TEM-based study on phage densities in six different soils (agricultural, coastal, and piedmont type) in Delaware, USA (Williamson et al., 2005) revealed VLP abundances of about 10^9 per g of soil. Other studies for example in field bulk soils at Cardiff and Oxford in the United Kingdom, showed similar VLP densities Thus, one may assume that, in most soils with bacteriomes of about 10^9 to 10^{10} cells per g, viral abundance is in this order, or above. However, soil virome data – with respect

to information on sequence types – are still fragmentary, as current datasets from soil ecosystems encompass only 10,009 viral contigs (IMG/VR)[i] (Paez-Espino et al., 2017a).

With respect to phages at plant roots, the rhizospheres and rhizosheaths of *Beta vulgaris* var. Amythest, *Poa pratensis* L., *Epibolium tetragonum* L. Griseb, *Senecio jacobaea* L., and *Cardamine flexuosa*, as well as from the surrounding bulk soils (fields in Cardiff and Oxford, UK), were investigated by TEM. Overall, counts were in the range 10^6–10^9 VLPs g^{-1} dry soil (Ashelford et al., 2003; Swanson et al., 2009). Remarkably, no differences in VLP (virion) abundances were found between the different rhizospheres, nor between rhizospheres and surrounding bulk soils (1.18 ± 0.014 x 10^9 VLPs g^{-1} dry soil for rhizosphere, and 1.17 ± 0.085 x 10^9 VLPs g^{-1} dry bulk soil) (Swanson et al., 2009). The rhizosheaths had 1.09 ± 0.014 x 10^9 VLPs g^{-1} dry weight. The lack of a difference between bulk and rhizosphere soils was awkward, as rhizospheres commonly contain raised bacterial densities. It was explained by (i) the physical or biochemical characteristics of the rhizosphere soil potentially limiting VLP counts, and/or (ii) bacteria selected in the rhizosphere potentially being less susceptible to virus infection. Here, we argue that the physicochemical nature of soil, in particular of the rhizosphere with its layered biofilm-like structure, indeed tends to promote the binding of viral particles, with consequently fewer VLPs being visualized than might be expected. Moreover, the 'more bacteria, less virions' (also known as **piggy-back-the-winner – PtW**) model may explain the findings (Knowles et al., 2016). The PtW strategy – which promotes lysogeny – has evolved to mitigate high host density conditions, as the energy cost of the superinfection immunity resulting from lysogenization is often lower than that resulting from mutation to phage resistance (Knowles et al., 2016; Touchon et al., 2017). The strategy is universally observed across environments, including soil settings (Knowles et al., 2016). Although the extent to which the PtW model drives phage–host interactions is still under debate (Weitz et al., 2017), we argue that it is, currently, a key model that may explain the often low VLP numbers observed in highly dense bacteriomes.

Viral diversity - marine versus soil ecosystems

Computational analyses of marine viromics data have, across a diversity of 867 viral types, clearly indicated that predicted tailed dsDNA bacteriophages are dominant in pelagic areas (Paez-Espino et al., 2017a; Roux et al., 2016a). In addition, ssDNA viruses have been reported in marine organisms, for example copepods (Dunlap et al., 2013) and the diatom *Chaetoceros setoensis* (Tomaru et al., 2013). The analyses also uncovered a plethora of previously unknown RNA viruses. For instance, (+)

ssRNA viruses were isolated from Jericho Pier (JP) and the Strait of Georgia (SOG), with the JP viruses being dominated by Picornavirales (identified as Coroviridae, Dicistroviridae, Marnaviridae, Picornaviridae, and Reoviridae) and the SOG viruses by Tombusviridae and the viral genus *Umbravirus* (Culley et al., 2006). Moreover, corals have been found to host a plethora of bacteriophages (identified on the basis of both morphological and sequence criteria), next to viruses of Eukarya (e.g., herpes, pox, and iridoviruses) (Thurber et al., 2017). We conclude that marine viral diversity is predictably high, rivalling organismal diversity. The specific setting in marine systems (e.g., pelagic versus benthic), driving organismal settlement and fate, is thought to also strongly determine viral diversity and role.

Our understanding of the diversity within soil viromes has so far been mostly built on TEM based morphological observations. For example, in rhizosphere and surrounding bulk soils, VLPs with tailed, spherical, rod-shaped, filamentous, or bacilliform morphologies have been found (Swanson et al., 2009). Several studies have now included nucleic-acid-based approaches and bioinformatics in their analyses of soil viromes (Adriaenssens et al., 2015; Scola et al., 2017; Srinivasiah et al., 2008, 2015; Wommack et al., 2015; Zablocki et al., 2014b, 2015). For instance, a study in desert soil showed the presence of a diversity of (predicted) tailed viruses (67% of the total; Myoviridae, Podoviridae and Siphoviridae). Moreover, the study uncovered novel Caudovirales (7%) and phages of *Geobacillus* (6%) and *Bacillus* (4%). Finally, a phage akin to deep-sea thermophilic phage D6E (1%), next to an ssDNA Microviridae phage (4%), was also found (Adriaenssens et al., 2015). Another computational survey of Antarctic soil viromes also revealed a high diversity of tailed phages (Zablocki et al., 2014b) as well as ssDNA phages belonging to the Inoviridae, Microviridae, Circoviridae, Geminiviridae, and Nanoviridae. Finally, several novel circular ssDNA viruses were discovered in paddy (Kim et al., 2008) and Antartic soils (Adriaenssens et al., 2017). Overall, based on the TEM and nucleic acid analyses, the virome diversities across these diverse soils were high and reflected the concomitant high species level bacterial diversities (Hatfull, 2015). What we do not yet understand is to what extent the viral diversity can be traced back to local events of organismal growth and evolution.

Ecological and evolutionary significance of bacterial viromes - marine versus soil ecosystems

As agents of HGT and reservoirs of ecologically relevant traits, phages can enhance the fitness of their hosts and also flexibly shape their evolution (Hurwitz et al., 2016). In particular circumstances, the host may even depend on its phages, in return securing phage reproduction (Obeng et al., 2016). How and why does this dependency

(or **coevolution**) occur? Coevolution arises or becomes more apparent as the reciprocal adaptation between interacting partners increases. However, what are the circumstances under which coevolution evolves? For instance, it is clearly observed in organisms that use counter-defence mechanisms. Thus, to avoid lysis by (virulent) phages, bacterial hosts may change their surface and/or internal phage-sensitive structures (Gomez and Buckling, 2013). The phage, in turn, may modify its anchoring or otherwise host-interactive device, enabling it to again productively interact with the host. An evolutionary 'arms race' has thus emerged. Overall, such (iterative) selective processes drive the diversity of the interacting partners, increasing evolutionary rates. Thus, host community structure is affected. Coevolution is dynamic and drives, in an iterative manner, the ecological fitness of the survivor (Koskella and Brockhurst, 2014). Here, a plethora of open questions arises. For instance, to what extent does the interaction between phages and their hosts affect the direct ecology of the partnership? In the marine system, this is strongly related to the lifestyles of the viruses. For instance, **virulent phages** serve as the controllers of host population abundances. When active, they are responsible for the release of **dissolved organic matter** (**DOM**) and **particulate organic matter** (**POM**) through a so-called **viral shunt** (Thurber et al., 2017) (**Figure 4.2**). The dynamics of population control exerted by virulent phages obeys the laws of classical models of predator–prey relationships. These are known as **kill-the-winner** (**KtW**) processes, in which the most dominant bacteria are, due to mass action rules, killed upon infection by virulent phages. Thus, a population balance of the attacked host, as well as a defined bacterial diversity, is maintained. KtW processes have been observed in both marine and soil ecosystems (Breitbart, 2012). The raised abundance of nutrients produced by the activity of phages on particular populations stimulate population-level as well as overall microbial community activity and nutrient cycling. As a consequence, evolution is accelerated and (parts of) the bacterial communities flourish. Such nutrient-releasing lysis events have been commonly observed in coral reef settings, with fundamental roles in maintaining the abundance and diversity of the local microbiomes (Thurber et al., 2017). The process is reminiscent of a phenomenon observed in bacterial biofilm formation. There, phage-induced host cell lysis [under conditions of high cell density, limited nutrient and oxygen availability, and accumulation of reactive oxygen species (ROS)] yields extracellular DNA (eDNA), which may become a structural part of the biofilm. The eDNA, as well as polysaccharides and additional nutrients, may then nurture the neighboring cells (Obeng et al., 2016; Secor et al., 2015) (**Figure 4.2**). KtW models reflect the events that occur in high-abundance host cell populations. In these, lytic infections are 'permitted', resulting in an increase in the **virus-to-bacteria ratio**. On the other hand, the competing PtW model may explain the course of events when the phage–host interaction results in lysogeny. Host-dependent cues are crucial for

the KtW versus PtW strategy. Possibly, intracellular 'decision systems', for example the phage 'arbitrium' system, are key to the choice of the lifestyle strategy, that is, virulent versus temperate (Erez et al., 2017).

Temperate phages (forming **prophages**) can affect the fitness of the host by lysogenic conversion. Such lysogenic conversions may be neutral, or tinker with host physiology, for example affecting host metabolic pathways, introducing so-called **auxiliary metabolic genes** (**AMGs**) (Breitbart, 2012; Hurwitz and U'Ren, 2016). Such AMGs may affect host metabolism during and/or after virus infection (Lindell et al., 2007; Roux et al., 2016a), for example changing host pathogenicity (Brüssow et al., 2004) or affecting host biofilm formation (Secor et al., 2015). Many phage-carried virulence factors underlying pathogenicity have earlier been identified as 'more-on (moron) DNA' (Brüssow et al., 2004) genes. Such virulence genes enable the host bacterium to broaden its environmental niche, and – in return – the phage optimizes its replication (Brüssow et al., 2004). The introduced traits affect the host cells at the individual level as well as at the population level (Obeng et al., 2016). Although a suite of morons has been uncovered, a majority of the predicted proteins is still unknown (Hurwitz et al., 2015). Consequently, a vast array of moron-encoded traits with potential to foster host fitness still awaits discovery.

To date, a large diversity of AMGs has been found in marine viromes. This ranges from genes encoding steps of different nutrient cycles, for example the sulfur (*dsr* and *sox*) and nitrogen cycles (*P-II* and *amoC*) (Roux et al., 2016a) to genes that directly enhance host-driven photosynthesis (Breitbart, 2012; Hurwitz and U'Ren, 2016; Lindell et al., 2005; Roux et al., 2016a). Concerning the latter, particular cyanophages carry photosynthesis genes (*psbA* and *psbD*, derived from *Prochlorococcus* and *Synechococcus*), enabling augmentation of host photosynthesis during infection. This process has great implications for global ocean productivity (Lindell et al., 2005, 2007). In reciprocity, the infected host, by its increased fitness in terms of energy capture, maximizes phage reproduction. Also, the aphotic (deep ocean) zone was found to be brimming with virus-carried AMGs for locally adaptive traits, such as those encoding motility proteins (*flaB*) and flagellar motor complexes (*motA*) (Hurwitz et al., 2015). In early work, performed in the 1970s, Reanney and coworkers already provided evidence for the potentially important roles of phages as population controllers in soil settings (Reanney and Marsh, 1973). Further, van Elsas and Pereira, (1987), in several studies on *Bacillus* phages in soil, confirmed these findings, pinpointing the potential key role of these phages in antibiotic resistance plasmid transduction. Such work was all based on prior cultivation of soil bacteria, and included TEM-based observations of their phages. Later work confirmed the finding of AMGs in soil viromes (Dinsdale et al., 2008). For example, the *phoH* gene (phosphate regulon gene) and a gene for ribonucleotide reductase (RNR) were found in a 'hypolithic' (below-

stone) virome (Adriaenssens et al., 2015). Phylogenetic analyses showed that the *phoH* genes (allocated to a soil cyanophage) were distantly related to similar genes from marine cyanophages (Goldsmith et al., 2011). Their prevalence in the hypolithic system suggested a significant function of *phoH*, for instance in phosphate acquisition (Adriaenssens et al., 2015). The high RNR gene abundance has not been reported for any other soil virome so far, but this gene might be advantageous in nutrient-limited environments (Adriaenssens et al., 2015). A study by Pal et al., (2007) showed that the evolutionary rate of soil-inhabiting *Pseudomonas fluorescens* SBW25 increases in the presence of a phage. However, a follow-up microcosm study contradicted this finding (Gomez and Buckling, 2013), as the authors found that the presence of virulent phages and a natural soil virome negatively affected the evolution of soil bacteria. Both studies addressed the effects of virulent phages as the agents driving bacterial evolution, failing to include temperate phages. Moreover, they did not take into account the heterogeneous nature of soil. Here, we argue that it is important to integrate both soil microhabitat/soil hotspot concepts and host– temperate phage interactions in the comprehensive framework of eco-evolutionary dynamics of host bacteria in the soil ecosystem (Obeng et al., 2016). Host–phage (long-term) coevolution is also presumed to foster the persistence of the rhizosphere bacterium *Pseudomonas putida* KT2440 in soil. Thus, in strain KT2440, 25 out of 105 genomic islands were found to relate to prophages, as well as transposons, insertion elements, and type II introns (Quesada et al., 2012). One prophage, denoted Pspu28, apparently modulated competitive fitness in the rhizosphere, as its excision and its lysogenic state both resulted in competitive fitness gains. However, the underlying mechanism is as yet unclear (Quesada et al., 2012). Thus, in spite of our current model of soil providing a unique compartmentalized 'evolutionary incubator' found in no other environment (Rillig et al., 2017) and the recently uncovered AMGs, the roles of most phages and AMGs in shaping the ecology and evolutionary trajectory of soil bacteria are still unclear.

Potential use of soil viruses

Interest in using the soil virome, especially phages, as plant disease biocontrol agents (phage therapy) dates back to the preantibiotic era. Unfortunately, our still-limited understanding of phage lifestyles in soil, at that time, has resulted in a greatly impaired development. The current state of soil/plant virome work, in connection to the need to develop biological techniques to treat plant diseases, has brought the attention back to the use of phages (Buttimer et al., 2017; Day et al., 2017). Phage therapy is already effective in the control of the plant pathogen *Dickeya solani*, which causes rot and blackleg in potato (Adriaenssens et al., 2012). In the light of the intricacies of the

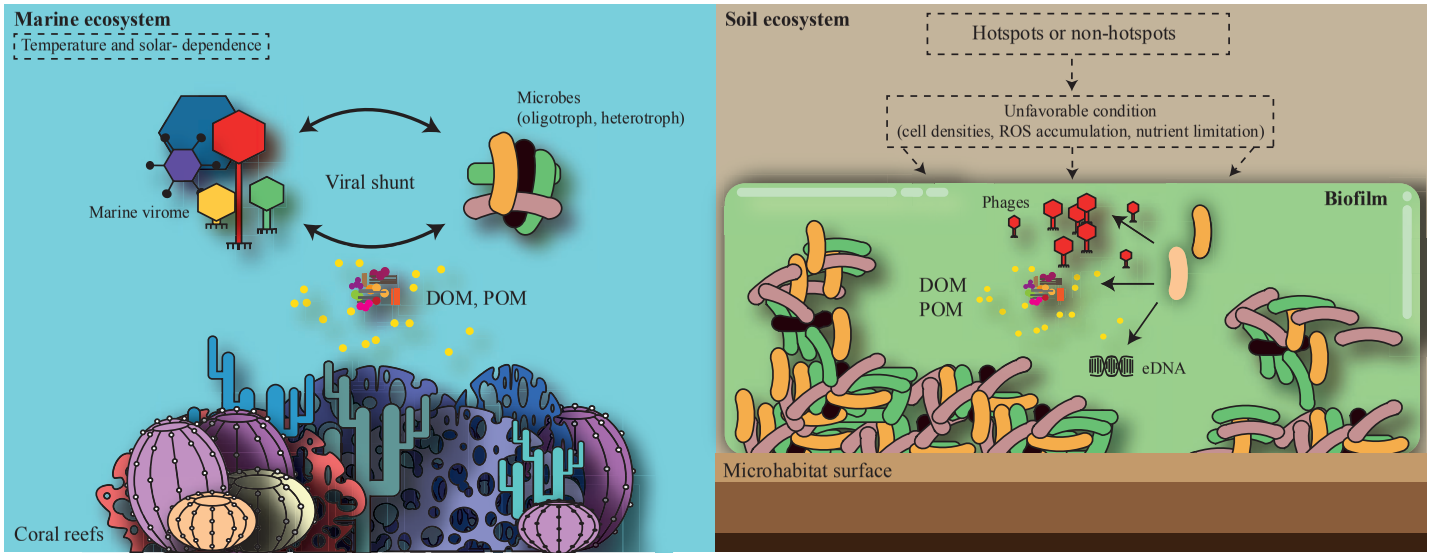

Figure 4.2. Example of virome dynamics in marine and soil ecosystems. In marine ecosystems the virome plays a significant role (Suttle, 2007). Marine viromes contribute to geochemical and nutrient cycles (Roux et al., 2016a), to organism mortality (Breitbart, 2012), and to modulation of marine communities. In the viral shunt process (Suttle, 2007), the marine virome controls the bacterial population in line with the predator–prey model known as 'kill-the-winner' (Breitbart, 2012). At the same time the lysis of bacteria by phages mediates nutrient release in the form of dissolved organic matter (DOM) and particulate organic matter (POM). On the other hand, biofilms, for example on the surface of a particle, can undergo the same process. Unfavorable conditions inside a biofilm, such as limited nutrients and oxygen, and accumulation of reactive oxygen species (ROS), lead to excision of phages; at the same time, the released materials (DOM, POM) serve as a 'public-good', providing readily-available nutrients to the neighboring cells. This process also provides extracellular chromosomal DNA (eDNA), which can form a structural part of the biofilm (Secor et al., 2015).

Table 4.1. Main technical challenges and state-of-the-art of soil virome research.

Technique	State-of-the-art	Soil sample application	Challenge(s)	References
Virome extraction	• Resuspention buffer: 1% potassium citrate (KC), amended 1% potassium citrate (AKC) and amended 5 mM sodium pyrophosphate (PP) • Physical dispersal: sonication, bead beating and vortexing • Sample storage condition: chill 4°Cand frozen -80°C	Broad soil type.	• Optimization in a different soil type is needed • No significant VLP count observed in any physical dispersal and sample storage condition • Combination of AKC, bead-beating/vortex yields higher VLPs	(Scola et al., 2017; Trubl et al., 2016; Williamson et al., 2013; Zablocki et al., 2014b)
Virome enumeration	• Transmission electron microscopy (TEM) • Epifluorescence microscopy (EFM) and • Flow cytometry/microscopy (FCM)	Broad sample origin.	• TEM gives [286TD$DIF]an approximation of VLP numbers (using internal standards) and types. Skills with TEM use required, time consuming. • EFM, FCM: skills required, including basic knowledge on viruses. • Fluorescence dye efficacy decreases due to aspecific interactions with soil materials.	(Ashelford et al., 2003; Kim et al., 2008; Swanson et al., 2009)
Virome concentration and purification	• Amicon filters with bovine serum albumin (BSA), polyethylene glycol (PEG) precipitation; sucrose and CsCl gradient centrifugations	Broad sample origin.	VLPs loss and reduced recovery	(Pratama and van Elsas, 2017; Trubl et al., 2016)
Virome amplification	• Multiple-displacement amplification (MDA) with ϕ29	Broad sample origin.	Preferable for ssDNA. This method is not recommended for dsDNA as it would cause undesirable background amplification/cause overrepresentation of virus composition	(Kim et al., 2008)

soil system, clear conditions of, for example, water content, pH, and salt level, are required to make phage-based biocontrol a success. Thus, if soil is water-saturated, the growth and movement of both bacteria and viruses – and thus their interactions – are promoted. Indeed, soils with high moisture content and rich in organic matter tend to have abundant viromes and bacteria (Srinivasiah et al., 2008). In the particular case of successful *D. solani* phage therapy, a suspension of 10^{10} PFU/l of phage LIMEstone1 was sprayed onto the potato tubers. Also, the presence of water films around the potato tubers originating from the phage (phages LIMEstone1 and LIMEstone2) application were speculated to promote the infection process.

Where should we go from here?

Soil viromics is clearly in its early stages compared to marine viromics. The reason is the difficulty, posed by soil, in examining phage–host interactive processes at the proper scale and level of detail. Indeed, we have hitherto gathered broad yet overall knowledge on soil viral abundances and diversities (Williamson et al., 2005). Moreover, some studies have addressed the effects of phages on particular bacterial populations (Adriaenssens et al., 2015; Scola et al., 2017; Zablocki et al., 2014a). However, an integrative perspective of how viruses shape the soil microbiome, and the components thereof, across the soil 'islands' where local events of adaptation take place, is not yet available. This is important, as one may consider the soil as a multiple parallel evolutionary incubator, with many processes of fitness modulations and consequent evolutionary steps taking place at the same time, even at microscale distances. Thus, one needs studies that are effective in the intricate analysis of the dynamics within soil viromes and the interactions with host organisms. Clearly, research efforts that match those in marine settings in magnitude are needed. However, the level of complexity is at least one step up in soil compared to marine settings, and this poses specific challenges to the relevant studies. The main challenges revolve around (*i*) the as yet underutilized technologies, both experimental (**Table 4**) and computational, *(ii*) the lack of knowledge about time- and site-specific phage life cycles and activities in soil, and consequentially, *(iii*) the lack of spatiotemporally explicit (microscale) sampling efforts, allowing the study of time courses of development. As a consequence, we currently lack a true understanding of the fine-scale, and robustly founded effects of the phage activities that drive local soil bacterial populations.

The proposed key approaches to the challenges are twofold. First, viral particles should be extracted from soil samples at fine scale, concentrated, purified, lysed, and sequenced. This approach enables assessment of the entire soil virome in one 'go', and needs to be miniaturized and performed in (high-throughput) parallel. One may

here capitalize on the rapid progress that has been made with respect to the technical approaches in marine viromics (Brum and Sullivan, 2015). However, for soil, the use of presumed or observed compartmentalization into microhabitat islands should be used as the guidance for soil dissection and sampling. Unfortunately, past optimized technical procedures have not yet yielded a universal protocol for this (Miura et al., 2011; Trubl et al., 2016; Williamson et al., 2005, 2013), as soils are widely divergent in texture, and technical challenges go beyond straight extraction and enumeration steps (Table 1). For instance, the choice of amplification method of extracted virome DNA and library preparation greatly affects the outcome of soil virome studies (Duhaime et al., 2012; Kim et al., 2008; Kim and Bae, 2011; Reavy et al., 2015; Roux et al., 2016b; Solonenko et al., 2013). Additionally, studies on the interactions between phages and bacterial hosts in the soil microhabitat pose great challenges (Vos et al., 2013). Recent studies in mycosphere hotspots (Zhang et al., 2014) showed **mobile genetic elements**, that is, IncP-1 and PromA plasmids, to be prevalent in the mycosphere. Moreover, the genome of the mycosphere inhabitant *Paraburkholderia terrae* BS001 contains a range of phage genes (Haq et al., 2014), whereas another *P. terrae* strain, BS437, was shown to release a novel inducible prophage, f437, spontaneously or following induction (Pratama and van Elsas, 2017). There is a need to further explore soil hotspots such as the mycosphere as to the extent and ecological significance of such accelerated parallel bacterial evolution trajectories; however, optimized approaches to this aim will still have to emerge.

The second approach is based on mining virome signals from soil bacterial metagenomes (Bolduc et al., 2017; Paez-Espino et al., 2017b; Roux et al., 2015). Unfortunately, metagenomics-based analyses easily miss viral particles, hence giving rise to incomplete viromes. Paez-Espino et al., (2016) mined 5 Tb of sequence data using publicly available viral sequences as references coupled with k-mer based viral binning, RNA polymerase (RNAp) tree domain analysis and extensive manual curation. They revealed numerous previously unknown viral groups in soil, next to freshwater, marine, thermal spring, plant, human, and engineered habitats, and also linked the viromes across ecosystems to their hosts. In this and other studies, virome metagenomics platforms, such as IMG/VR, were important, as they provide a unique integration of large virome meta-data (260 000 viral sequences from >6000 ecological origins, both viral isolate reference and virome metagenomics) and analytic tools (Paez-Espino et al., 2017a). Other bottlenecks in viromics studies pertain to the current viromics platforms placing little emphasis on non-dsDNA viruses. For example, methods to address ssDNA viruses (usually giving short sequences: <5–10 kb), RNA viruses, and giant viruses, are still under development. Moreover, viral sequences constituting only a very small fraction of bacterial

metagenomic sequences, reveal uneven distributions across habitats, differences in sequence length, in numbers of reads or samples studied, and hence pose additional computational hurdles. Moreover, the high inter- and intra-population viral diversities might cause ambiguities in sequence assemblies (Rose et al., 2016). There is, thus, the need for well-developed viral metagenomics sequencing as well as sequence analysis platforms, with sufficient sequence coverage (Rose et al., 2016). Fortunately, key platforms have been recently developed, that is, the publicly accessible VIROME (Wommack et al., 2012), Metavir (Roux et al., 2014), VirSorter (Roux et al., 2015) and iVirus (Bolduc et al., 2017) platforms. These not only mine virome signals from bacterial metagenomics data but also analyze these viromics data.

Apart from the general lack of spatiotemporally explicit data in most soil viromics studies, there is a perceived lack of studies across diverse soil types (**Figure 4.3**).

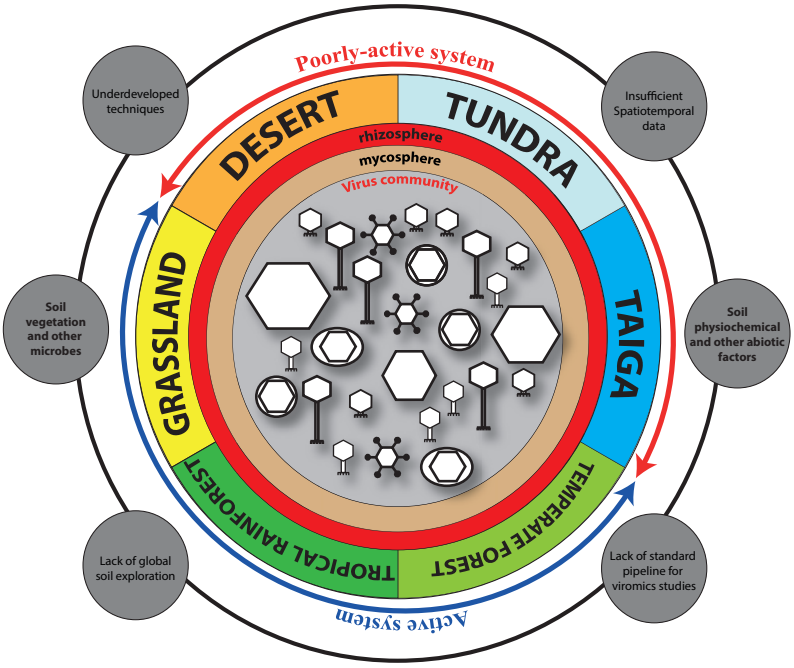

Figure 4.3. The extent of the role of viromes in soil has yet to be discovered. The challenges that soil virome studies face are mainly due to the still underutilized techniques (Table 1) and an insufficient focus on spatiotemporal data. There is also the need for global soil explorations, as soil is explicitly diverse, such as tundra, taiga, temperate forest, tropical rainforest, grassland, and desert. The characteristic nature of each soil habitat affects the local bacterial communities (Fierer et al., 2012). There is a need for a standardized pipeline to mine soil virome signals and analyse these. The bottleneck in soil viromics is related to the physicochemical nature of soil, as well as to other abiotic and biotic factors.

Globally, the major soil habitats include (*i*) tundra, (*ii*) boreal forest, (*iii*) temperate forest, (*iv*) tropical rainforest, (*v*) grassland, and (*vi*) desert. The sheer diversity across and within these soil habitats (microscale diversity) is in line with differences in soil bacterial diversity (Fierer et al., 2012). Yet, how such diversity affects the structure and diversity of the corresponding soil viromes is still an open question. Clearly, the current scarcity of directed efforts to explore the viromics of broad soil ecosystems at the microscale level is tightly connected with the technical limitations (**Table 4**) and the nature of the soil.

Concluding remarks and future perspectives

In this opinion paper we addressed the knowledge gap between our understanding of the viromes of marine versus soil ecosystems. Marine virology has recently developed fast and has started to incorporate viral function into ecosystem models. Soil viromics will need to develop in a similar way. There is a sense of urgency here, as soil viruses – in particular phages – play fundamental roles in the ecologies of their hosts and so influence key soil processes. Future studies need to consider the spatiotemporal aspects and the connectivity bottleneck of the soil habitat, resulting in the concept of compartmentalization (see Outstanding Questions). Thus, the island-like patches in soil, with their separate evolutionary trajectories, need to be placed in the spotlight. The extent to which such trajectories are shared across the islands, or are different, is a potential focus for research. Moreover, tailored technologies for viral detection and observation need to be optimized in order to overcome the barriers posed by the nature of soil. In particular, we need to address the phage–host interaction events in soil hotspots such as the mycosphere (Zhang et al., 2014) and the rhizosphere (van Elsas et al., 2003). In particular, the litter layers of (rain) forests may contain highly diverse active microorganisms that are prone to time-dependent interactions with phages. The required spatial exploration of soil ecosystems can only be addressed with temporal variability dynamics studies. Soil chronosequence studies have already shown changes in bacterial community structures and bacterial trait evolution, with potential involvement of phages (Dini-Andreote et al., 2014). Furthermore, the study of soil viromes is important to foster our understanding of the fate of human pathogens (e.g. *E. coli* and *Salmonella* spp. entering the food chain from soil into crop plants). Finally, to obtain a meaningful integrative perspective of the role of viromes in soil microbiomes, we advocate that hypothesis-based rather than broadly descriptive studies are performed.

Acknowledgments

This work was supported through a scholarship of the Indonesia Endowment Fund for Education to (LPDP - Lembaga Pengelola Dana Pendidikan, Departemen Keuangan, Republik Indonesia) A.A.P. Further, we would like to thank Dr. Stefanie Vink and Yustina Yusuf for a helpful discussion. We would also like to thank you Dr Gail Teitzel for structural comments on an early draft and, importantly, two anonymous reviewers for their critical and helpful comments.

Outstanding Questions

How to overcome the technical bottlenecks that impair the study of the soil virome?

Do soil viromes have higher abundance and diversity compared to marine viromes? What drives this abundance and diversity? Is there any 'generalist' virome across the soil ecosystems?

How different are soil virome structures across different spatiotemporal scales (long-term/short-term and different soil ecosystems, as well as different soil depth)?

To what extent does the soil virome affect the ecological functions of the soil?

How can the soil virome contribute to the parallel evolutionary trajectories inside the soil aggregates?

Can long-term spatiotemporal observations (e.g., chronosequences) answer to what extent the soil virome affects the changes in bacterial community structures and bacterial trait
evolution?

Resources

[i]https://img.jgi.doe.gov/cgi-bin/vr/main.cgi (accessed on October 20, 2017).

Keukenhof, 2015